Defence Research and Development Canada
Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

A simple expression for the matrix gradient of a diagonal element of R in QR decomposition

for use in MIMO Communications and Signal Processing

A. Yasotharan

Defence R&D Canada – Ottawa

Technical Memorandum
DRDC Ottawa TM 2010-247
December 2010

Canada

A simple expression for the matrix gradient of a diagonal element of R in QR decomposition

for use in MIMO Communications and Signal Processing

A. Yasotharan

Defence R&D Canada – Ottawa

Technical Memorandum

DRDC Ottawa TM 2010-247

December 2010

Author

Original signed by A. Yasotharan

A. Yasotharan

Approved by

Original signed by Bill Katsube

Bill Katsube
Head/Communications and Navigation Electronic Warfare

Approved for release by

Original signed by Chris McMillan

Chris McMillan
Head/Document Review Panel

Abstract

The QR matrix decomposition (QRD), or factorization, has many applications. In matrix computations, it is used to solve linear equations and least squares problems. In signal processing, it is used for adaptive filtering, adaptive beamforming/ interference nulling, and direction finding. In communications, it is used for adaptive equalization and transceiver design for multiple-input multiple-output (MIMO) channels.

When QRD is used in signal processing and communications, it is of interest to know the effects of noise. This work is a first step towards that goal.

Given matrix $\mathbf{A}$ with full column-rank M, we consider the unique decomposition $\mathbf{A} = \mathbf{QR}$ where $\mathbf{Q}$ is a matrix with M orthonormal columns and $\mathbf{R}$ is an $M \times M$ upper triangular matrix with real positive diagonal elements $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M$. Treating $\hat{r}_i$ as a function of the elements of $\mathbf{A}$, a simple expression is derived for its matrix gradient with respect to $\mathbf{A}$.

Future work will aim to derive expressions for the gradients of all elements of $\mathbf{Q}$ and $\mathbf{R}$, and use these expressions to evaluate the effect of noise perturbation in $\mathbf{A}$. The present result is useful in optimizing certain MIMO decision-feedback communication systems.

Résumé

La décomposition QR (QRD) d'une matrice, ou sa factorisation, a plusieurs applications. Dans les calculs matriciels, elle permet de résoudre des équations linéaires et des problèmes des moindres carrés. Dans le traitement de signaux, elle sert au filtrage adaptatif, à la mise en forme adaptative de faisceaux, à la suppression de brouillage et à la radiogoniométrie. Dans les communications, elle sert à l'égalisation adaptative et à la conception d'émetteurs récepteurs avec canaux MIMO (entrées multiples, sorties multiples).

Lorsque la décomposition QR est utilisée dans le traitement de signaux et les communications, il est important de connaître les effets du bruit. Les présents travaux constituent la première étape vers la réalisation de cet objectif.

Soit une matrice $\mathbf{A}$ de rang-colonne complet M, nous considérons la décomposition unique $\mathbf{A} = \mathbf{QR}$ où $\mathbf{Q}$ est une matrice M à colonnes orthonormales et $\mathbf{R}$ est une matrice triangulaire supérieure $M \times M$ dont les éléments diagonaux sont des nombres réels positifs $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M$. Considérant $\hat{r}_i$ comme une fonction des éléments $\mathbf{A}$, une expression simple est dérivée de son gradient matriciel par rapport à $\mathbf{A}$.

Les recherches futures auront comme objectif de dériver des expressions pour les gradients des éléments de $\mathbf{Q}$ et de $\mathbf{R}$, et d'utiliser ces expressions pour évaluer les effets de la perturbation sur le bruit dans la matrice $\mathbf{A}$. Les résultats de la présente étude seront utiles dans l'optimisation de certains systèmes de communication à décision rétroactive MIMO.

This page intentionally left blank.

Executive summary

A simple expression for the matrix gradient of a diagonal element of R in QR decomposition

A. Yasotharan; DRDC Ottawa TM 2010-247; Defence R&D Canada – Ottawa; December 2010.

Background: In linear algebra, it is well known that any matrix $\mathbf{A}$ can be decomposed, or factorized, as $\mathbf{A} = \mathbf{Q}\mathbf{R}$ where $\mathbf{Q}$ is a matrix with orthonormal columns and $\mathbf{R}$ is an upper triangular matrix. This so-called QR decomposition (QRD) has many applications. In matrix computations, it is used to solve linear equations and least squares problems. In signal processing, it is used for adaptive filtering, adaptive beamforming/ interference nulling, and direction finding. In communications, it is used for adaptive equalization and transceiver design for multiple-input multiple-output (MIMO) channels.

In signal processing and communications, it is of interest to know how noise perturbation of $\mathbf{A}$ affects the elements of $\mathbf{Q}$ and $\mathbf{R}$. This work is a first step towards that goal.

Principal results: In general the QRD is not unique. When $\mathbf{A}$ has full column-rank M, uniqueness can be ensured by choosing the matrix $\mathbf{Q}$ to have M columns and constraining the diagonal elements of $\mathbf{R}$, say $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M$, to be positive. We consider this case.

Treating $\hat{r}_i$ as a function of the elements of $\mathbf{A}$, a simple expression is derived for its matrix gradient with respect to $\mathbf{A}$. Let $\mathbf{A} = \mathbf{X} + j\mathbf{Y}$ be the real-imaginary decomposition of $\mathbf{A}$. Let $\mathbf{q}_i$ and $\mathbf{s}_i$ be the i^{th} columns of $\mathbf{Q}$ and $\mathbf{S} = \mathbf{R}^{-1}$ respectively. It is shown that $\frac{\partial}{\partial \mathbf{A}}\hat{r}_i = \frac{\partial}{\partial \mathbf{X}}\hat{r}_i + j\frac{\partial}{\partial \mathbf{Y}}\hat{r}_i = \hat{r}_i \mathbf{q}_i \mathbf{s}_i^H$, which is a rank-one matrix.

Matrix gradients of functions of $\{\hat{r}_i\}$ of the form $f(\hat{r}_i)$ and $f(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M)$ are obtained.

Significance of results: When $\mathbf{A}$ is a signal-plus-noise matrix whose QRD is sought, it is of interest to know how noise affects $\mathbf{Q}$ and $\mathbf{R}$. Here, the gradients of elements of $\mathbf{Q}$ and $\mathbf{R}$ with respect to $\mathbf{A}$ will be of value, as seen in stochastic perturbation theory. This work constitutes a first step in that direction.

In MIMO decision-feedback communication system design, the QRD of the transmitter-channel composite matrix yields the optimum receiver and its performance. Specifically, $\{\hat{r}_i^2\}$ are the signal-to-noise ratios for the parallel data streams. Thus the derived expressions for $\frac{\partial}{\partial \mathbf{A}}\hat{r}_i$ should be useful in jointly optimizing the transmitter and the receiver.

Future work: Expressions must be derived for the gradients of all elements of $\mathbf{Q}$ and $\mathbf{R}$, as these expressions will be useful, according to the stochastic perturbation theory, in evaluating the effects of noise in the QRD of a signal-plus-noise matrix.

It would be worthwhile to demonstrate that the derived expressions for $\frac{\partial}{\partial \mathbf{A}}\hat{r}_i$ are useful in MIMO system design.

Sommaire

A simple expression for the matrix gradient of a diagonal element of R in QR decomposition

A. Yasotharan ; DRDC Ottawa TM 2010-247 ; R & D pour la défense Canada – Ottawa ; décembre 2010.

Introduction : En algèbre linéaire, il est bien connu que n'importe quelle matrice $\mathbf{A}$ peut être décomposée, ou factorisée, puisque $\mathbf{A} = \mathbf{QR}$ où $\mathbf{Q}$ est une matrice à colonnes orthonormales et $\mathbf{R}$ est une matrice triangulaire supérieure. Cette décomposition QR (QRD) a plusieurs applications. Dans les calculs matriciels, elle permet de résoudre des équations linéaires et des problèmes des moindres carrés. Dans le traitement de signaux, elle sert au filtrage adaptatif, à la mise en forme adaptative de faisceaux, à la suppression de brouillage et à la radiogoniométrie. Dans les communications, elle sert à l'égalisation adaptative et à la conception d'émetteurs récepteurs avec canaux MIMO (entrées multiples, sorties multiples).

Pour le traitement du signal et les communications, il est important de connaître les effets de la perturbation sur le bruit de la matrice $\mathbf{A}$ sur les éléments de $\mathbf{Q}$ et $\mathbf{R}$. Les présents travaux constituent la première étape vers la réalisation de cet objectif.

Résultats : En général, la décomposition QR n'est pas unique. Lorsque $\mathbf{A}$ est à rang-colonne complet M, on peut garantir l'unicité en donnant M colonnes à la matrice $\mathbf{Q}$ et en contraignant les éléments diagonaux de $\mathbf{R}$, disons $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M$, à être positifs. Nous considérons ce cas.

En considérant $\hat{r}_i$ comme une fonction des éléments de $\mathbf{A}$, une expression simple est dérivée de son gradient matriciel par rapport à $\mathbf{A}$. Soit $\mathbf{A} = \mathbf{X} + j\mathbf{Y}$ la décomposition réel/imaginaire de $\mathbf{A}$. Soit $\mathbf{q}_i$ et $\mathbf{s}_i$ les i^{e} colonnes de $\mathbf{Q}$ et $\mathbf{S} = \mathbf{R}^{-1}$ respectivement. On a $\frac{\partial}{\partial \mathbf{A}}\hat{r}_i = \frac{\partial}{\partial \mathbf{X}}\hat{r}_i + j\frac{\partial}{\partial \mathbf{Y}}\hat{r}_i = \hat{r}_i \mathbf{q}_i \mathbf{s}_i^H$, qui est une matrice de rang un.

Les gradients matriciels des fonctions de $\{\hat{r}_i\}$ de la forme $f(\hat{r}_i)$ et $f(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M)$ sont obtenus.

Portée : Lorsque l'on veut calculer la décomposition QR d'une matrice signal plus bruit $\mathbf{A}$, il est intéressant de connaître les effets du bruit sur $\mathbf{Q}$ et $\mathbf{R}$. À cette fin, les gradients des éléments $\mathbf{Q}$ et $\mathbf{R}$ par rapport à $\mathbf{A}$ seront importants, comme vu dans la théorie de la perturbation stochastique. Les présents travaux constituent la première étape vers la réalisation de cet objectif.

Dans la conception d'un système de communication à décision rétroactive MIMO, La décomposition QR de la matrice composite du canal émetteur donne un récepteur dont le rendement est optimal. Plus précisément, $\left\{\hat{r}_i^2\right\}$ sont les rapports signal-bruit des flux de données parallèles. Les expressions dérivées pour $\frac{\partial}{\partial \mathbf{A}}\hat{r}_i$ devraient donc être utiles pour optimiser conjointement l'émetteur et le récepteur.

Recherches futures : Des expressions doivent être dérivées pour les gradients des éléments de $\mathbf{Q}$ et $\mathbf{R}$, car ces expressions seront utiles, selon la théorie de perturbation stochastique, pour évaluer les effets du bruit dans la décomposition QR d'une matrice signal plus bruit.

Il serait bon d'établir la preuve que les expressions dérivées pour $\frac{\partial}{\partial \mathbf{A}}\hat{r}_i$ sont utiles à la conception d'un système MIMO.

Table of contents

Acknowledgements

Gerard Peiris (CARDS) reviewed this document and made many comments that helped to improve the presentation. Sichun Wang (CNEW) reviewed a preliminary version and made useful comments.

1 Introduction

In linear algebra, it is well known that any matrix $\mathbf{A}$ can be decomposed, or factorized, as $\mathbf{A} = \mathbf{QR}$ where $\mathbf{Q}$ is a matrix with orthonormal columns and $\mathbf{R}$ is an upper triangular matrix. A formal statement of this result is given in Section 2. This so-called QR decomposition has many applications in matrix computations - solving linear equations and linear least squares fitting problems, to name a couple [1].

The QR decomposition also has many applications in signal processing and communications. The signal processing applications include adaptive filtering [2] (chapters 14,15), adaptive beamforming [2] (chapter 14.5) [3] (chapter 7), and direction-finding [4]. The adaptive beamforming methods that can be efficiently implemented via QR decomposition include LSE and MVDR/MPDR [3]. These methods are also applicable in interference nulling, eg. GPS anti-jamming. The communications applications include adaptive equalization [2] (chapter 15.9) and transceiver design for multiple-input multiple-output (MIMO) communications [5] [6] [7].

In the above signal processing and communications applications, the QR decomposition of a signal-plus-noise matrix is found. Therefore, it is of interest to know how the noise affects the factors $\mathbf{Q}$ and $\mathbf{R}$. If the noise variance is small enough, the effect of noise can be studied via the Stochastic Perturbation Theory [8]. A brief overview of this theory is as follows. Let A be a matrix and let F be a matrix-valued function of A with derivative $F_A^{'}$. Given a perturbation matrix E, presumed small, we can write

$$F(A+E) \approx F(A) + F_A^{'}(E). \tag{1}$$

Suppose E is random and of the cross-correlated type defined in [8]. Then the difference

$$F(A+E) - F(A) \approx F_A^{'}(E) \tag{2}$$

can be estimated in a stochastic sense, provided we have a tractable expression for the derivative $F_A^{'}(E)$.

In the context of the QR decomposition $\mathbf{A} = \mathbf{QR}$, we want to study how the factors $\mathbf{Q}$ and $\mathbf{R}$ are affected when $\mathbf{A}$ is perturbed by noise. The work reported herein constitutes a first step towards that goal. Treating the diagonal elements $\{\hat{r}_i\}$ of $\mathbf{R}$ as functions of $\mathbf{A}$, we obtain a simple generic expression for their gradients with respect to $\mathbf{A}$; since we are dealing with scalar-valued functions, gradient and derivative are the same. Future work will aim to derive expressions for the gradients of all elements of $\mathbf{Q}$ and $\mathbf{R}$. These gradients may then be used to study how noise in matrix $\mathbf{A}$ affects it QRD.

The rest of the report is organized as follows:

Section 2 provides the needed mathematical preliminaries: a theorem on the *QR decomposition*, the definition of the *matrix gradient*, and notations. It also states the scope of this report. Section 3 states a result concerning the orthogonal projection of a vector onto the column-span of a matrix. This result is used in Section 4 to derive a simple generic expression for the gradients of $\left\{\hat{r}_i^2\right\}$. The main result of the report - a simple expression for the gradients of $\{\hat{r}_i\}$ - is given in Section 4 along with other results. Section 5 outlines an application of the result to MIMO communication system design. Section 6 summarizes the results and suggests some lines of further work. Annexes A and B derive gradients that are used in Section 3.

2 Mathematical Preliminaries and Scope

2.1 QR Decomposition

We begin with a well known result in Linear Algebra. See [9] (Theorem 2.6.1) and [10] (Section 5.2, especially, Section 5.2.6).

Theorem 1 (QR decomposition) *Suppose $\mathcal{A}$ is a $K \times M$ complex matrix with $K \geq M$ and rank$(\mathcal{A}) = M$. Then $\mathcal{A}$ can be factored as*

$$\mathcal{A} = \mathcal{Q}\mathcal{R} \tag{3}$$

where $\mathcal{Q}$ is a $K \times M$ matrix with orthonormal columns, i.e., $\mathcal{Q}^H \mathcal{Q} = \mathbf{I}_M$, the $M \times M$ identity matrix, and $\mathcal{R}$ is an $M \times M$ upper triangular nonsingular matrix.

If we insist that the diagonal elements of $\mathcal{R}$, say $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M$, are positive real, then $\mathcal{Q}$ and $\mathcal{R}$ are unique. □

The above factorization of $\mathcal{A}$ is the so-called *thin* QR factorization defined in [10] (text preceding Theorem 5.2.2 in page 230). Its uniqueness when $\mathcal{R}$ has positive diagonal elements is asserted by Theorem 5.2.2 and Section 5.2.10 of [10]. In terms of this unique factorization, any general QR decomposition will have the form $\mathcal{A} = \tilde{\mathcal{Q}}\tilde{\mathcal{R}}$ where $\tilde{\mathcal{Q}} = \mathcal{Q}\mathcal{P}$ and $\tilde{\mathcal{R}} = \mathcal{P}^*\mathcal{R}$ for some diagonal matrix $\mathcal{P}$ with diagonal elements of unit absolute value. Here * denotes complex conjugation.

Given a matrix, its QR decomposition can be computed by several methods, e.g. Gram-Schmidt orthogonalization, Householder reflections, Givens rotations, etc [10] (Section 5.2). The method that is the most relevant to the present paper is the Gram-Schmidt orthogonalization. The properties of QR decomposition given in [10] (Theorem 5.2.1) are also very useful.

2.2 Matrix gradient

Definition 1 (Matrix gradient) *Let $\mathbf{A}$ be a complex-valued matrix and let $\mathbf{A} = \mathbf{X} + j\mathbf{Y}$ be its real-imaginary decomposition. For a real-valued scalar function c of $\mathbf{A}$, we denote by $\frac{\partial c}{\partial \mathbf{X}}$ the matrix of partial derivatives of c with respect to the elements of $\mathbf{X}$. Similarly, we denote by $\frac{\partial c}{\partial \mathbf{Y}}$ the matrix of partial derivatives of c with respect to the elements of $\mathbf{Y}$. Thus $\frac{\partial c}{\partial \mathbf{X}}$ and $\frac{\partial c}{\partial \mathbf{Y}}$ are real-valued matrices of the same size as $\mathbf{A}$. We also denote*

$$\frac{\partial c}{\partial \mathbf{A}} = \frac{\partial c}{\partial \mathbf{X}} + j\frac{\partial c}{\partial \mathbf{Y}}. \tag{4}$$

We call $\frac{\partial c}{\partial \mathbf{X}}$ the matrix gradient *of c with respect to $\mathbf{X}$, and similarly for $\frac{\partial c}{\partial \mathbf{Y}}$. We call $\frac{\partial c}{\partial \mathbf{A}}$ the* matrix gradient *of c with respect to $\mathbf{A}$.* □

Note that when $\mathbf{A}$ is a vector, the *matrix gradient* reduces to the *vector gradient*.

We use the term 'gradient' to refer to both vector and matrix gradients, as the notation will indicate which type of gradient is being referred to.

There are many books and papers that discuss the gradient of a real-valued scalar function of a vector or matrix variable. A well known paper is [11]. A well known book is [12]. A more mathematical treatment is given in [13], the preface of which gives an extensive bibliography on the theory and applications of matrix differential calculus. Appendix A.7 of [3] and the appendix of Chapter 6 of [14] give gradients of some commonly encountered scalar functions of vectors and matrices. All of the above, except [3], consider only real-valued scalar functions of real-valued vector or matrix variables. The gradient of a real-valued scalar function of a complex-valued vector is discussed in [3] (A.7.4) with acknowledgement to [15].

2.3 Notation

For a complex matrix $\mathbf{A}$, we denote by $\mathrm{ran}(\mathbf{A})$ the *range* of $\mathbf{A}$, ie. the vector subspace spanned by the columns of $\mathbf{A}$ [10](Section 2.1.2). The orthogonal complement of $\mathrm{ran}(\mathbf{A})$ is denoted by $\mathrm{ran}^{\perp}(\mathbf{A})$.

2.4 Scope of Report

In the general context of the unique QR decomposition of Theorem 1 above, first we derive a simple expression for $\frac{\partial}{\partial \mathcal{A}} \hat{r}_i^2$. From this, we derive $\frac{\partial}{\partial \mathcal{A}} \hat{r}_i$, and more generally, $\frac{\partial}{\partial \mathcal{A}} f(\hat{r}_i)$ for any differentiable $f(.)$. As an example of the latter, we derive $\frac{\partial}{\partial \mathcal{A}} \log \hat{r}_i$, from which we derive $\frac{\partial}{\partial \mathcal{A}} \log \det(\mathcal{R})$ which is equivalent to $\frac{1}{2} \frac{\partial}{\partial \mathcal{A}} \log \det(\mathcal{A}^H \mathcal{A})$. We also treat the general case $\frac{\partial}{\partial \mathcal{A}} f(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M)$ for $f(.)$ that has all partial derivatives.

3 Gradient of Squared Norm of Orthogonal Projection

Let $\mathbf{A}$ be a $K \times L$ ($K \geq L$) matrix with rank L and let $\mathbf{a}$ be a $K \times 1$ vector. In this section, we will derive a result concerning the projection of $\mathbf{a}$ onto $\text{ran}^{\perp}(\mathbf{A})$. In the next section, we will apply this result to the QR dcomposition of (3) by considering $\mathbf{A}$ to be the matrix of the first L columns of $\mathcal{A}$ and $\mathbf{a}$ to be the $(L+1)^{\text{th}}$ column of $\mathcal{A}$, for $L = 1, 2, \ldots, (M-1)$.

The orthogonal projection matrix $\mathbf{P}$ onto $\text{ran}^{\perp}(\mathbf{A})$ is

$$\mathbf{P} = \mathbf{I}_K - \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H \tag{5}$$

which is Hermitian ($\mathbf{P}^H = \mathbf{P}$) and idempotent ($\mathbf{P}^2 = \mathbf{P}$). Denoting by $\hat{r}$ the norm of the orthogonal projection of $\mathbf{a}$ onto $\text{ran}^{\perp}(\mathbf{A})$, we have

$$\begin{aligned}
\hat{r}^2 &= \|\mathbf{Pa}\|^2 & (6)\\
&= \mathbf{a}^H\mathbf{P}^H\mathbf{Pa} & (7)\\
&= \mathbf{a}^H\mathbf{Pa}. & (8)
\end{aligned}$$

First we will derive the gradients of $\hat{r}^2$ with respect to $\mathbf{a}$ and $\mathbf{A}$, as expressions involving $\mathbf{a}$ and $\mathbf{A}$. Then we will simplify the expressions using the QR decomposition of the augmented matrix $[\mathbf{A}, \mathbf{a}]$.

We will use the term ‘gradient’ for vector and matrix gradients when the context is clear.

The gradient of $\hat{r}^2$ with respect to $\mathbf{a}$ is (see Annex A, Eq. (A.1))

$$\frac{\partial}{\partial \mathbf{a}}\hat{r}^2 = 2\mathbf{Pa}. \tag{9}$$

The gradient of $\hat{r}^2$ with respect to $\mathbf{A}$ is given by the following lemma whose statement makes use of the pseudo-inverse of $\mathbf{A}$ given by

$$\mathbf{A}^{\#} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H. \tag{10}$$

Lemma 1

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{A}}\hat{r}^2 &= -2\left(\mathbf{I}_K - \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\right)\mathbf{a}\mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1} & (11)\\
&= -2\left(\mathbf{Pa}\right)\left(\mathbf{A}^{\#}\mathbf{a}\right)^H. & (12)
\end{aligned}$$

Proof: See Annex B for the proof of (11). Then use (5) and (10) to get (12). $\square$

Note that when $\mathbf{a} \in \mathrm{ran}(\mathbf{A})$, we have $\frac{\partial}{\partial \mathbf{A}} \hat{r}^2 = \mathbf{0}$. To avoid this, we assume henceforth that $\mathbf{a} \notin \mathrm{ran}(\mathbf{A})$. Then $\frac{\partial}{\partial \mathbf{A}} \hat{r}^2$ is a rank-one matrix since $(\mathbf{P}\mathbf{a})$ is a column vector and $\left(\mathbf{A}^{\#}\mathbf{a}\right)^H$ is a row vector.

In order to simplify (9) and (12), suppose $\mathbf{A}$ and $[\mathbf{A}, \mathbf{a}]$ have the QR decompositions

$$\mathbf{A} = \mathbf{Q}\mathbf{R} \tag{13}$$

and

$$[\mathbf{A}, \mathbf{a}] = \tilde{\mathbf{Q}}\tilde{\mathbf{R}}. \tag{14}$$

which are unique in the sense of Theorem 1.

Then $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{R}}$ can be partitioned as [10] (Theorem 5.2.1)

$$\tilde{\mathbf{Q}} = [\mathbf{Q}, \mathbf{q}] \tag{15}$$

$$\tilde{\mathbf{R}} = \begin{bmatrix} \mathbf{R} & \mathbf{r} \\ \mathbf{0} & \hat{r} \end{bmatrix} \tag{16}$$

where $\mathbf{q}$ is a $K \times 1$ vector, $\mathbf{r}$ is a $L \times 1$ vector and $\hat{r}$ is a real-valued scalar which is equal to the norm of the projection of $\mathbf{a}$ onto $\mathrm{ran}^{\perp}(\mathbf{A})$ defined in (6). The latter fact can be seen as follows. Eqs. (14), (15), and (16) together show

$$\mathbf{a} = \mathbf{Q}\mathbf{r} + \mathbf{q}\hat{r} \tag{17}$$

which is an orthogonal decomposition because

$$\mathbf{Q}^H \mathbf{q} = \mathbf{0}. \tag{18}$$

Moreover, since $\mathbf{Q}\mathbf{r} \in \mathrm{ran}(\mathbf{A})$, we have

$$\mathbf{q}\hat{r} = \mathbf{P}\mathbf{a} \tag{19}$$

which together with $\|\mathbf{q}\| = 1$ shows that $\hat{r}$ is the norm of the projection of $\mathbf{a}$ onto $\mathrm{ran}^{\perp}(\mathbf{A})$.

Using (19), the gradient of (9) can be written simply as

$$\frac{\partial}{\partial \mathbf{a}} \hat{r}^2 = 2\mathbf{q}\hat{r}. \tag{20}$$

To simplify (12), we need some extra algebra. Denote $\mathbf{S} = \mathbf{R}^{-1}$. Then $\mathbf{S}$ is upper triangular [10] (Section 3.1.8, The Algebra of Triangular Matrices). Moreover, $\tilde{\mathbf{R}}^{-1}$, which is also upper triangular, can be partitioned as [9] (Section 0.7.3)

$$\tilde{\mathbf{R}}^{-1} = \begin{bmatrix} \mathbf{R} & \mathbf{r} \\ \mathbf{0} & \hat{r} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{S} & \mathbf{s} \\ \mathbf{0} & \hat{s} \end{bmatrix} \tag{21}$$

where $\mathbf{s}$ is a $L \times 1$ vector and $\hat{s}$ is a scalar given by

$$\begin{aligned} \mathbf{R}^{-1}\mathbf{r} &= -\hat{r}\mathbf{s} && (22) \\ \hat{r}\hat{s} &= 1. && (23) \end{aligned}$$

The following lemma gives a simple expression for the matrix gradient of (12).

Lemma 2

$$\frac{\partial}{\partial \mathbf{A}} \hat{r}^2 = 2\hat{r}^2 \mathbf{q}\mathbf{s}^H. \tag{24}$$

Proof:

Using (13), the pseudo-inverse of (10) can be written as $\mathbf{A}^{\#} = \mathbf{R}^{-1}\mathbf{Q}^H$. Using this and (17), we get

$$\begin{aligned} \mathbf{A}^{\#}\mathbf{a} &= \mathbf{R}^{-1}\mathbf{Q}^H(\mathbf{Q}\mathbf{r} + \mathbf{q}\hat{r}) && (25) \\ &= \mathbf{R}^{-1}\mathbf{r} && (26) \\ &= -\hat{r}\mathbf{s}. && (27) \end{aligned}$$

where in the last step we have used (22). Using (27) and (19) in (12), we get (24). □

By combining (24) and (20), we obtain the following result.

Lemma 3

$$\frac{\partial}{\partial [\mathbf{A}, \mathbf{a}]} \hat{r}^2 = 2\hat{r}^2 \mathbf{q}[\mathbf{s}^H, \hat{s}]. \tag{28}$$

Proof: Using (23), (20) can be written as $\frac{\partial}{\partial \mathbf{a}} \hat{r}^2 = 2\hat{r}^2 \mathbf{q}\hat{s}$. Combine this with (24) □

Note that $[\mathbf{s}^H, \hat{s}]$ is the Hermitian transpose of the last column of $\tilde{\mathbf{R}}^{-1}$ whose partitioned form is given by (21).

4 The Main Result

Referring to Theorem 1 and Eq. (3), we first introduce the notation needed for stating our main result. Denote by $\mathbf{q}_i$ the i^{th} column of $\mathcal{Q}$. Denote $\mathcal{S} = \mathcal{R}^{-1}$ and denote by $\mathbf{s}_i$ the i^{th} column of $\mathcal{S}$. Thus

$$\mathcal{Q} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M) \tag{29}$$
$$\mathcal{R}^{-1} = \mathcal{S} = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M). \tag{30}$$

Note that $\mathcal{S}$ is also upper triangular [10] (Section 3.1.8). Recall that $\{\hat{r}_i\}$ are the diagonal elements of $\mathcal{R}$.

The following theorem is the main result of this paper.

Theorem 2 *Referring to (3) and the above notations,*

$$\frac{\partial}{\partial \mathcal{A}} \hat{r}_i^2 = 2\hat{r}_i^2 \mathbf{q}_i \mathbf{s}_i^H \qquad \textit{for } i = 1, 2, \ldots, M. \tag{31}$$

$\square$

The proof rests on the fact that QR decomposition can be done via Gram-Schmidt orthogonalization. Denote by $\mathbf{a}_i$ the i^{th} column of $\mathcal{A}$ of (3). Denote $\mathbf{A}_i = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_i]$ for $i = 1, 2, \ldots, M$. Then $\mathbf{a}_1 = \mathbf{q}_1 \hat{r}_1$ and, for $i > 1$, $\mathbf{q}_i \hat{r}_i$ is the orthogonal projection of $\mathbf{a}_i$ onto $\text{ran}^{\perp}(\mathbf{A}_{i-1})$. Evidently, $\hat{r}_i^2$ does not depend on $[\mathbf{a}_{i+1}, \mathbf{a}_{i+2}, \ldots, \mathbf{a}_M]$, and we have

$$\frac{\partial}{\partial \mathcal{A}} \hat{r}_i^2 = \left[\frac{\partial}{\partial \mathbf{A}_i} \hat{r}_i^2, \qquad \mathbf{0}_{K \times (M-i)} \right]. \tag{32}$$

Denote by $\{\hat{s}_i : i = 1, 2, \ldots, M\}$ the diagonal elements of $\mathbf{S}$. Then (cf. (23))

$$\hat{s}_i \hat{r}_i = 1 \qquad \text{for } i = 1, 2, \ldots, M. \tag{33}$$

The cases $i = 1$ and $i > 1$ are separately treated below.

Proof for Case $i = 1$

Observe that

$$\hat{r}_1^2 = \mathbf{a}_1^H \mathbf{a}_1 \tag{34}$$
$$\mathbf{a}_1 = \mathbf{q}_1 \hat{r}_1. \tag{35}$$

Therefore,

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{A}_1}\hat{r}_1^2 &= \frac{\partial}{\partial \mathbf{a}_1}(\mathbf{a}_1^H\mathbf{a}_1) && (36)\\
&= 2\mathbf{a}_1 && (37)\\
&= 2\mathbf{q}_1\hat{r}_1 && (38)\\
&= 2\mathbf{q}_1\hat{r}_1^2\hat{s}_1 && (39)
\end{aligned}
$$

where (37) follows from (A.1) of Annex A and in the last step we have used (33). Combining this with (32), we get

$$
\begin{aligned}
\frac{\partial}{\partial \mathcal{A}}\hat{r}_1^2 &= 2\mathbf{q}_1\hat{r}_1^2[\hat{s}_1, \mathbf{0}_{1\times(M-1)}] && (40)\\
&= 2\mathbf{q}_1\hat{r}_1^2\mathbf{s}_1^H && (41)
\end{aligned}
$$

where in the last step we have used the fact that $\mathcal{S}$ of (30) is upper triangular. □

Proof for Case $i > 1$

Denote by $\mathbf{R}_i$ the leading $i \times i$ submatrix of $\mathcal{R}$, the upper triangular matrix of (3). Define the partition

$$
\mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{i-1} & \mathbf{r}_i \\ \mathbf{0} & \hat{r}_i \end{bmatrix} \qquad (42)
$$

for $i = 2, 3, \ldots, M$, where we identify $\mathbf{R}_1 = \hat{r}_1$. Denote by $\mathbf{S}_i$ the leading $i \times i$ submatrix of $\mathcal{S}$, the upper triangular matrix of (30). Define the partition

$$
\mathbf{S}_i = \begin{bmatrix} \mathbf{S}_{i-1} & \tilde{\mathbf{s}}_i \\ \mathbf{0} & \hat{s}_i \end{bmatrix} \qquad (43)
$$

for $i = 2, 3, \ldots, M$, where we identify $\mathbf{S}_1 = \hat{s}_1$. Then $\mathbf{S}_i\mathbf{R}_i = \mathbf{I}_i$ for $i = 1, 2, \ldots, M$ (cf. (21)). Denote $\mathbf{Q}_i = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_i]$ for $i = 1, 2, \ldots, M$.

Since $\hat{r}_i$ is the norm of the orthogonal projection of $\mathbf{a}_i$ onto $\mathrm{ran}^{\perp}(\mathbf{A}_{i-1})$, we invoke Lemma 3 of Section 3 by setting $\mathbf{A} = \mathbf{A}_{i-1}$, $\mathbf{a} = \mathbf{a}_i$, $\mathbf{Q} = \mathbf{Q}_{i-1}$, $\mathbf{q} = \mathbf{q}_i$, $\mathbf{R} = \mathbf{R}_{i-1}$, $\mathbf{r} = \mathbf{r}_i$, $\hat{r} = \hat{r}_i$, $\mathbf{S} = \mathbf{S}_{i-1}$, $\mathbf{s} = \tilde{\mathbf{s}}_i$, $\hat{s} = \hat{s}_i$, to get

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{A}_i}\hat{r}_i^2 &= \frac{\partial}{\partial [\mathbf{A}_{i-1}, \mathbf{a}_i]}\hat{r}_i^2 && (44)\\
&= 2\mathbf{q}_i\hat{r}_i^2\left[\tilde{\mathbf{s}}_i^H, \hat{s}_i\right]. && (45)
\end{aligned}
$$

Observe that $[\tilde{\mathbf{s}}_i^H, \hat{s}_i]$ is the i^{th} row of $\mathbf{S}_i^H$. Combining this with (32), we get

$$
\begin{aligned}
\frac{\partial}{\partial \mathcal{A}}\hat{r}_i^2 &= 2\mathbf{q}_i\hat{r}_i^2\left[\tilde{\mathbf{s}}_i^H, \hat{s}_i, \mathbf{0}_{1\times(M-i)}\right] && (46)\\
&= 2\mathbf{q}_i\hat{r}_i^2\mathbf{s}_i^H && (47)
\end{aligned}
$$

where in the last step we have used the fact that $\mathcal{S}$ of (30) is upper triangular. $\square$

The following corollary gives the gradient of $\hat{r}_i$. This result is the subject of the title of this paper.

Corollary 1

$$\frac{\partial}{\partial \mathcal{A}} \hat{r}_i = \hat{r}_i \mathbf{q}_i \mathbf{s}_i^H. \tag{48}$$

Proof: By the chain rule of differentiation,

$$\begin{aligned} \frac{\partial}{\partial \mathcal{A}} \hat{r}_i &= \left(\frac{\partial}{\partial (\hat{r}_i^2)} \hat{r}_i \right) \left(\frac{\partial}{\partial \mathcal{A}} \hat{r}_i^2 \right) & (49) \\ &= \left(\frac{1}{2\hat{r}_i} \right) \left(2\hat{r}_i^2 \mathbf{q}_i \mathbf{s}_i^H \right) & (50) \\ &= \hat{r}_i \mathbf{q}_i \mathbf{s}_i^H. & (51) \end{aligned}$$

$\square$

More generally, the gradient of any differentiable function of $\hat{r}_i$ is given by the following corollary.

Corollary 2 *Let the real-valued function $f(x)$ be differentiable in $(0, \infty)$ and let $f'(x)$ be its derivative. Then*

$$\frac{\partial}{\partial \mathcal{A}} f(\hat{r}_i) = f'(\hat{r}_i) \hat{r}_i \mathbf{q}_i \mathbf{s}_i^H. \tag{52}$$

$\square$

As an example, we have

$$\frac{\partial}{\partial \mathcal{A}} \log \hat{r}_i = \mathbf{q}_i \mathbf{s}_i^H. \tag{53}$$

Weighted linear combinations over $i = 1, 2, \ldots, M$ can be taken of the above gradients. Let $\{\theta_i : i = 1, 2, \ldots, M\}$ be a set of weights, and denote the diagonal matrix

$$\boldsymbol{\Theta} = \operatorname{diag}(\theta_1, \theta_2, \ldots, \theta_M). \tag{54}$$

Denote the diagonal matrix

$$\hat{\mathbf{R}} = \operatorname{diag}(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M) = \operatorname{diag}(\mathcal{R}). \tag{55}$$

By taking weighted sums of (31), we get

$$\frac{\partial}{\partial \mathcal{A}}\left(\sum_{i=1}^{M}\theta_i \hat{r}_i^2\right) = \sum_{i=1}^{M} 2\theta_i \hat{r}_i^2 \mathbf{q}_i \mathbf{s}_i^H \tag{56}$$

$$\tag{57}$$

$$= 2\mathcal{Q}\boldsymbol{\Theta}\hat{\mathbf{R}}^2\mathcal{S}^H. \tag{58}$$

Similarly, by taking weighted sums of (48), we get

$$\frac{\partial}{\partial \mathcal{A}}\left(\sum_{i=1}^{M}\theta_i \hat{r}_i\right) = \mathcal{Q}\boldsymbol{\Theta}\hat{\mathbf{R}}\mathcal{S}^H. \tag{59}$$

By directly summing (53) over $i = 1, 2, \ldots, M$, we get

$$\frac{\partial}{\partial \mathcal{A}}\log\det(\mathcal{R}) = \frac{\partial}{\partial \mathcal{A}}\log\prod_{i=1}^{M}\hat{r}_i \tag{60}$$

$$= \frac{\partial}{\partial \mathcal{A}}\sum_{i=1}^{M}\log\hat{r}_i \tag{61}$$

$$= \mathcal{Q}\mathcal{S}^H. \tag{62}$$

This can also be written as

$$\frac{\partial}{\partial \mathcal{A}}\log\det(\mathcal{A}^H\mathcal{A}) = 2\mathcal{Q}\mathcal{S}^H \tag{63}$$

since $\mathcal{A}^H\mathcal{A} = \mathcal{R}^H\mathcal{R}$ and $\det(\mathcal{A}^H\mathcal{A}) = \det(\mathcal{R})^2$.

For a square ($K = M$) and nonsingular $\mathbf{A}$, (62) reduces to

$$\frac{\partial}{\partial \mathcal{A}}\log|\det(\mathcal{A})| = \mathcal{A}^{-H}. \tag{64}$$

More generally, let $f(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M)$ be a real-valued function, and denote

$$\hat{\mathbf{F}} = \operatorname{diag}\left(\frac{\partial f}{\partial \hat{r}_1}, \frac{\partial f}{\partial \hat{r}_2}, \ldots, \frac{\partial f}{\partial \hat{r}_M}\right). \tag{65}$$

Then, by the chain rule of differentiation,

$$\frac{\partial}{\partial \mathcal{A}} f(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M) = \sum_{i=1}^{M}\frac{\partial f}{\partial \hat{r}_i}\frac{\partial}{\partial \mathcal{A}}\hat{r}_i \tag{66}$$

$$= \sum_{i=1}^{M}\frac{\partial f}{\partial \hat{r}_i}\left(\hat{r}_i \mathbf{q}_i \mathbf{s}_i^H\right) \tag{67}$$

$$= \mathcal{Q}\hat{\mathbf{F}}\hat{\mathbf{R}}\mathcal{S}^H. \tag{68}$$

4.1 The Real Case

When the given matrix $\mathcal{A}$ of (3) is real, its unique QR decomposition $\mathcal{A} = \mathcal{Q}\mathcal{R}$ yields real matrices $\mathcal{Q}$ and $\mathcal{R}$. Therefore, the above expressions for gradients can be used with the *Hermitian* transpose being interpreted as the *normal* transpose. Thus

$$\frac{\partial}{\partial \mathbf{A}} \hat{r}_i = \hat{r}_i \mathbf{q}_i \mathbf{s}_i^T . \tag{69}$$

5 Application to MIMO Communications

In multiple-input-multiple-output (MIMO) decision-feedback (DF) communication system theory, the optimum receiver for a given transmitter and channel can be described easily via the QR decomposition of the transmitter-channel composite matrix [6] [5] [7]. Moreover, when the optimum receiver is used, the values $\{\hat{r}_i^2\}$ represent the Signal-to-Noise Ratios (SNRs) for the parallel data streams that are being communicated.

Problems of optimizing the transmitter-receiver pair of a MIMO DF system generally fall into two categories:

1. maximize performance as measured by a function $f(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M)$ subject to a constraint on the transmitter power
2. minimize the transmitter power subject to constraints on $\{\hat{r}_i\}$ or functions thereof.

To derive first-order optimality conditions for these problems, we need an expression for the matrix gradient of $f(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M)$ or $\{\hat{r}_i\}$ with respect to the elements of the transmitter matrix. Such an expression can be obtained from (48) and the chain rule of differentiation; note that (48) gives the gradient w.r.t. the transmitter-channel composite matrix.

In fact, an expression for the matrix gradient of a weighted sum of the SNRs was derived in [7] (Appendix A) and used to optimize the transmitter and Zero-Forcing (ZF) receiver of a MIMO DF system. That derivation directly deals with the QR decomposition of the transmitter-channel composite matrix, and therefore is complicated. Moreover, it hides the general results that would apply in other situations, eg. a MIMO DF system with a Minimum Mean Square Error (MMSE) receiver. The derivations of the present report are not only much simpler but also expose the general results. Using the results presented herein, gradients of general performance measures, w.r.t. the transmitter matrix, can be derived easily for both ZF and MMSE receivers.

6 Conclusion

6.1 Summary

For a full-column-rank matrix $\mathcal{A}$, we considered its unique QR decomposition $\mathcal{A} = \mathcal{Q}\mathcal{R}$ such that $\mathcal{Q}^H\mathcal{Q} = I$ and $\mathcal{R}$ is square upper triangular with positive diagonal elements, say $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M$.

Treating $\hat{r}_i$ as a function of the elements of $\mathcal{A}$, we derived a simple expression for the matrix gradient of $\hat{r}_i^2$ (Theorem 2). We then combined this with the chain rule of differentiation to obtain simple expressions for the matrix gradients of $\hat{r}_i$ (Corollary 1) and a general differentiable real-valued function $f(\hat{r}_i)$ (Corollary 2). As an example of the latter, we derived the gradient of $\log \hat{r}_i$ (53). Using this we derived the gradient of $\log\det(\mathcal{A}^H\mathcal{A})$ (63). By combining Corollary 1 with the chain rule, we also derived the gradient of a general real-valued differentiable function $f(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M)$ (68).

We noted how the main result of this report may be used to optimize the transmitter-receiver pair of a MIMO DF communication system.

6.2 Suggestions for Future Work

Expressions must be derived for the gradients of all elements of $\mathcal{Q}$ and $\mathcal{R}$, as these will be useful in evaluating the effects of noise in $\mathcal{A}$, according to the stochastic perturbation theory.

It would be worthwhile to demonstrate the utility of the results of this report in optimizing the transmitter-receiver pair of a MIMO DF communication system under various criteria.

Annex A: Gradient of a Quadratic Form

For a complex-valued vector $\mathbf{a}$ and a Hermitian-symmetric matrix $\mathbf{B}$, consider the real-valued quadratic form $\mathbf{a}^H\mathbf{B}\mathbf{a}$. Here we derive the gradient

$$\frac{\partial}{\partial \mathbf{a}}\left(\mathbf{a}^H\mathbf{B}\mathbf{a}\right) = 2\mathbf{B}\mathbf{a}. \tag{A.1}$$

Although this is a simple result, its derivation will be a good introduction to the somewhat complicated derivations in Annex B. This result is used to get (9) and (37).

For any generic real-valued scalar variable α, which can be the real or imaginary part of any element of $\mathbf{a}$, we can write, by the product rule of differentiation,

$$\frac{\partial}{\partial \alpha}\left(\mathbf{a}^H\mathbf{B}\mathbf{a}\right) = \left(\frac{\partial \mathbf{a}^H}{\partial \alpha}\right)\mathbf{B}\mathbf{a} + \mathbf{a}^H\mathbf{B}\left(\frac{\partial \mathbf{a}}{\partial \alpha}\right). \tag{A.2}$$

Let $\mathbf{a} = \mathbf{x} + j\mathbf{y}$ be the real-imaginary decomposition of $\mathbf{a}$. Let x_n be the n^{th} element of $\mathbf{x}$ and similarly for y_n. For a general vector z, we denote by $z|_n$ the n^{th} element of z.

We shall use the following facts below. For $\alpha = x_n$, $\frac{\partial \mathbf{a}}{\partial \alpha}$ is a vector that has 1 at position n and zeros elsewhere. Similarly, for $\alpha = y_n$, $\frac{\partial \mathbf{a}}{\partial \alpha}$ is a vector that has j at position n and zeros elsewhere.

Setting $\alpha = x_n$, we have by (A.2)

$$\begin{aligned} \frac{\partial}{\partial x_n}\left(\mathbf{a}^H\mathbf{B}\mathbf{a}\right) &= \left.(\mathbf{B}\mathbf{a})\right|_n + \left.(\mathbf{B}\mathbf{a})\right|_n^* && \text{(A.3)} \\ &= 2\Re\left.(\mathbf{B}\mathbf{a})\right|_n. && \text{(A.4)} \end{aligned}$$

Letting α vary over $\mathbf{x}$, we have

$$\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{a}^H\mathbf{B}\mathbf{a}\right) = 2\Re\left(\mathbf{B}\mathbf{a}\right). \tag{A.5}$$

Setting $\alpha = y_n$, we have by (A.2)

$$\begin{aligned} \frac{\partial}{\partial y_n}\left(\mathbf{a}^H\mathbf{B}\mathbf{a}\right) &= -j\left.(\mathbf{B}\mathbf{a})\right|_n + j\left.(\mathbf{B}\mathbf{a})\right|_n^* && \text{(A.6)} \\ &= 2\Im\left.(\mathbf{B}\mathbf{a})\right|_n. && \text{(A.7)} \end{aligned}$$

Letting α vary over $\mathbf{y}$, we have

$$\frac{\partial}{\partial \mathbf{y}}\left(\mathbf{a}^H\mathbf{B}\mathbf{a}\right) = 2\Im\left(\mathbf{B}\mathbf{a}\right). \tag{A.8}$$

By combining (A.5) and (A.8) according to Definition 1 of Section 2, we get (A.1).

This page intentionally left blank.

Annex B: Proof of Lemma 1 of Section 3

This annex proves (11) of Lemma 1.

From (5) and (8), we have

$$\hat{r}^2 = \mathbf{a}^H\mathbf{a} - \mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{a}. \tag{B.1}$$

The gradient, w.r.t. $\mathbf{A}$, of the first term of (B.1) is zero. The gradient of the second term of (B.1) can be obtained by the product rule of differentiation as follows. For any generic real-valued scalar variable α, which can be the real or imaginary part of any element of $\mathbf{A}$, we can write

$$\frac{\partial}{\partial\alpha}\left(\mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{a}\right) = d_1(\alpha) + d_2(\alpha) + d_3(\alpha) \tag{B.2}$$

where

$$\begin{aligned} d_1(\alpha) &= \mathbf{a}^H\left(\frac{\partial\mathbf{A}}{\partial\alpha}\right)(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{a} && \text{(B.3)}\\ d_2(\alpha) &= \mathbf{a}^H\mathbf{A}\left(\frac{\partial(\mathbf{A}^H\mathbf{A})^{-1}}{\partial\alpha}\right)\mathbf{A}^H\mathbf{a} && \text{(B.4)}\\ d_3(\alpha) &= \mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\left(\frac{\partial\mathbf{A}^H}{\partial\alpha}\right)\mathbf{a}. && \text{(B.5)} \end{aligned}$$

Let $\mathbf{A} = \mathbf{X} + j\mathbf{Y}$ be the real-imaginary decomposition of $\mathbf{A}$. In the following subsections, we shall evaluate $d_1(\alpha)$, $d_2(\alpha)$, and $d_3(\alpha)$ when α varies over $\mathbf{X}$ and $\mathbf{Y}$. Towards this, we denote

$$\begin{aligned} \mathbf{b} &= (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{a} && \text{(B.6)}\\ \mathbf{c} &= \mathbf{A}\mathbf{b}. && \text{(B.7)} \end{aligned}$$

We shall use the following facts below. For $\alpha = x_{m,n}$, $\frac{\partial\mathbf{A}}{\partial\alpha}$ is a matrix that has 1 at position (m,n) and zeros elsewhere. Similarly, for $\alpha = y_{m,n}$, $\frac{\partial\mathbf{A}}{\partial\alpha}$ is a matrix that has j at position (m,n) and zeros elsewhere.

Notation

For a vector z, we denote by $z|_n$ the n^{th} element of z.

For a matrix Z, we denote by $Z|_{m,n}$ the $(m,n)^{\text{th}}$ element of Z.

d_1 over X and Y

Using (B.6) in (B.3), we have

$$d_1(\alpha) = \mathbf{a}^H \left(\frac{\partial \mathbf{A}}{\partial \alpha} \right) \mathbf{b}. \tag{B.8}$$

Setting $\alpha = x_{m,n}$, we have

$$\begin{aligned} d_1(x_{m,n}) &= \mathbf{a}|_m^* \, \mathbf{b}|_n & \text{(B.9)} \\ &= (\mathbf{a}\mathbf{b}^H)\big|_{m,n}^* . & \text{(B.10)} \end{aligned}$$

Setting $\alpha = y_{m,n}$, we have

$$\begin{aligned} d_1(y_{m,n}) &= j \, \mathbf{a}|_m^* \, \mathbf{b}|_n & \text{(B.11)} \\ &= j \, (\mathbf{a}\mathbf{b}^H)\big|_{m,n}^* . & \text{(B.12)} \end{aligned}$$

d_3 over X and Y

Using (B.6) in (B.5), we have

$$d_3(\alpha) = \mathbf{b}^H \left(\frac{\partial \mathbf{A}^H}{\partial \alpha} \right) \mathbf{a}. \tag{B.13}$$

Setting $\alpha = x_{m,n}$, we have

$$\begin{aligned} d_3(x_{m,n}) &= \mathbf{b}|_n^* \, \mathbf{a}|_m & \text{(B.14)} \\ &= (\mathbf{a}\mathbf{b}^H)\big|_{m,n} . & \text{(B.15)} \end{aligned}$$

Setting $\alpha = y_{m,n}$, we have

$$\begin{aligned} d_3(y_{m,n}) &= -j \, \mathbf{b}|_n^* \, \mathbf{a}|_m & \text{(B.16)} \\ &= -j \, (\mathbf{a}\mathbf{b}^H)\big|_{m,n} . & \text{(B.17)} \end{aligned}$$

$d_1 + d_3$ over X and Y

Combining (B.10) and (B.15), we get

$$\begin{aligned} d_1(x_{m,n}) + d_3(x_{m,n}) &= (\mathbf{a}\mathbf{b}^H)\big|_{m,n}^* + (\mathbf{a}\mathbf{b}^H)\big|_{m,n} & \text{(B.18)} \\ &= 2\Re\left((\mathbf{a}\mathbf{b}^H)\big|_{m,n} \right). & \text{(B.19)} \end{aligned}$$

Combining (B.12) and (B.17), we get

$$d_1(y_{m,n}) + d_3(y_{m,n}) = j\,(\mathbf{a}\mathbf{b}^H)\big|_{m,n}^{*} - j\,(\mathbf{a}\mathbf{b}^H)\big|_{m,n} \tag{B.20}$$

$$= 2\Im\left((\mathbf{a}\mathbf{b}^H)\big|_{m,n} \right). \tag{B.21}$$

Simplification of $d_2(\alpha)$

Suppose $\mathbf{Z}(\alpha)$ is a non-singular matrix-valued function of the real-valued scalar α. Then, by applying the product rule of differentiation to the identity

$$\mathbf{Z}(\alpha)\mathbf{Z}^{-1}(\alpha) = \mathbf{I} \tag{B.22}$$

and rearranging the terms, we get the well known result

$$\frac{\partial \mathbf{Z}^{-1}}{\partial \alpha} = -\mathbf{Z}^{-1}(\alpha)\left(\frac{\partial \mathbf{Z}}{\partial \alpha}\right)\mathbf{Z}^{-1}(\alpha). \tag{B.23}$$

Setting $\mathbf{Z} = (\mathbf{A}^H\mathbf{A})$, we get

$$\frac{\partial (\mathbf{A}^H\mathbf{A})^{-1}}{\partial \alpha} = -(\mathbf{A}^H\mathbf{A})^{-1}\left(\frac{\partial (\mathbf{A}^H\mathbf{A})}{\partial \alpha}\right)(\mathbf{A}^H\mathbf{A})^{-1}. \tag{B.24}$$

By applying the product rule to the middle term, we get

$$\frac{\partial (\mathbf{A}^H\mathbf{A})^{-1}}{\partial \alpha} = -(\mathbf{A}^H\mathbf{A})^{-1}\left[\left(\frac{\partial \mathbf{A}^H}{\partial \alpha}\right)\mathbf{A} + \mathbf{A}^H\left(\frac{\partial \mathbf{A}}{\partial \alpha}\right)\right](\mathbf{A}^H\mathbf{A})^{-1} \tag{B.25}$$

Using the above in (B.4), we get

$$d_2(\alpha) = -\mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\left[\left(\frac{\partial \mathbf{A}^H}{\partial \alpha}\right)\mathbf{A} + \mathbf{A}^H\left(\frac{\partial \mathbf{A}}{\partial \alpha}\right)\right](\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{a}. \tag{B.26}$$

Using (B.6) in the above, we get

$$d_2(\alpha) = -\mathbf{b}^H\left[\left(\frac{\partial \mathbf{A}^H}{\partial \alpha}\right)\mathbf{A} + \mathbf{A}^H\left(\frac{\partial \mathbf{A}}{\partial \alpha}\right)\right]\mathbf{b}. \tag{B.27}$$

Using (B.7) in the above, and changing sign, we get

$$-d_2(\alpha) = \mathbf{b}^H\left(\frac{\partial \mathbf{A}^H}{\partial \alpha}\right)\mathbf{c} + \mathbf{c}^H\left(\frac{\partial \mathbf{A}}{\partial \alpha}\right)\mathbf{b}. \tag{B.28}$$

d_2 over X and Y

Note that the first term of the right-hand side of (B.28) is similar to $d_3(\alpha)$ of (B.13) and the second term is similar to $d_1(\alpha)$ of (B.8).

Setting $\alpha = x_{m,n}$, we have

$$\begin{aligned} -d_2(x_{m,n}) &= \left.(\mathbf{c}\mathbf{b}^H)\right|_{m,n} + \left.(\mathbf{c}\mathbf{b}^H)\right|^*_{m,n} & \text{(B.29)} \\ &= 2\Re\left(\left.(\mathbf{c}\mathbf{b}^H)\right|_{m,n}\right). & \text{(B.30)} \end{aligned}$$

Setting $\alpha = y_{m,n}$, we have

$$\begin{aligned} -d_2(y_{m,n}) &= -j\left.(\mathbf{c}\mathbf{b}^H)\right|_{m,n} + j\left.(\mathbf{c}\mathbf{b}^H)\right|^*_{m,n} & \text{(B.31)} \\ &= 2\Im\left(\left.(\mathbf{c}\mathbf{b}^H)\right|_{m,n}\right). & \text{(B.32)} \end{aligned}$$

$d_1 + d_2 + d_3$ over X and Y

Combining (B.19) and (B.30), we get

$$\begin{aligned} d_1(x_{m,n}) + d_2(x_{m,n}) + d_3(x_{m,n}) &= 2\Re\left(\left.(\mathbf{a}\mathbf{b}^H)\right|_{m,n}\right) - 2\Re\left(\left.(\mathbf{c}\mathbf{b}^H)\right|_{m,n}\right) & \text{(B.33)} \\ &= 2\Re\left(\left.(\mathbf{a}\mathbf{b}^H - \mathbf{c}\mathbf{b}^H)\right|_{m,n}\right) & \text{(B.34)} \\ &= 2\Re\left(\left.((\mathbf{a}-\mathbf{c})\mathbf{b}^H)\right|_{m,n}\right). & \text{(B.35)} \end{aligned}$$

Combining (B.21) and (B.32), we get

$$\begin{aligned} d_1(y_{m,n}) + d_2(y_{m,n}) + d_3(y_{m,n}) &= 2\Im\left(\left.(\mathbf{a}\mathbf{b}^H)\right|_{m,n}\right) - 2\Im\left(\left.(\mathbf{c}\mathbf{b}^H)\right|_{m,n}\right) & \text{(B.36)} \\ &= 2\Im\left(\left.(\mathbf{a}\mathbf{b}^H - \mathbf{c}\mathbf{b}^H)\right|_{m,n}\right) & \text{(B.37)} \\ &= 2\Im\left(\left.((\mathbf{a}-\mathbf{c})\mathbf{b}^H)\right|_{m,n}\right). & \text{(B.38)} \end{aligned}$$

Gradient of Second Term of (B.1)

From (B.2) and (B.35), we get

$$\frac{\partial}{\partial \mathbf{X}}\left(\mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{a}\right) = 2\Re\left((\mathbf{a}-\mathbf{c})\mathbf{b}^H\right). \qquad \text{(B.39)}$$

From (B.2) and (B.38), we get

$$\frac{\partial}{\partial \mathbf{Y}}\left(\mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{a}\right) = 2\Im\left((\mathbf{a}-\mathbf{c})\mathbf{b}^H\right). \tag{B.40}$$

By combining the above, according to Definition 1 of Section 2, we write

$$\frac{\partial}{\partial \mathbf{A}}\left(\mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{a}\right) = 2(\mathbf{a}-\mathbf{c})\mathbf{b}^H. \tag{B.41}$$

Expansion of $(\mathbf{a}-\mathbf{c})\mathbf{b}^H$

By using (B.7) and (B.6), we can expand $(\mathbf{a}-\mathbf{c})\mathbf{b}^H$ in terms of $\mathbf{a}$ and $\mathbf{A}$ as follows:

$$\begin{aligned}
(\mathbf{a}-\mathbf{c})\mathbf{b}^H &= (\mathbf{a}-\mathbf{A}\mathbf{b})\mathbf{b}^H && \text{(B.42)}\\
&= (\mathbf{a}-\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{a})\mathbf{b}^H && \text{(B.43)}\\
&= (\mathbf{I}_K-\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H)\mathbf{a}\mathbf{b}^H && \text{(B.44)}\\
&= (\mathbf{I}_K-\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H)\mathbf{a}\mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}. && \text{(B.45)}
\end{aligned}$$

Final Results

From (B.1), (B.41), and (B.45), we get

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{A}}\hat{r}^2 &= -\frac{\partial}{\partial \mathbf{A}}\left(\mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{a}\right) && \text{(B.46)}\\
&= -2(\mathbf{a}-\mathbf{c})\mathbf{b}^H && \text{(B.47)}\\
&= -2\left(\mathbf{I}_K-\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\right)\mathbf{a}\mathbf{a}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1} && \text{(B.48)}
\end{aligned}$$

which is (11) of Lemma 1.

References

[1] N. J. Higham, Accuracy and Stability of Numerical Algorithms, 2nd ed., SIAM, 2002.

[2] S. Haykin, Adaptive Filter Theory, Prentice-Hall, 3rd ed., 1996.

[3] H. L. Van Trees, Optimum Array Processing - Part IV of Detection Estimation and Modulation Theory, Wiley-Interscience, 2002.

[4] C. H. Bischof and G. M. Shroff, On Updating Signal Subspaces, IEEE Transactions on Signal Processing, Vol. 40, NO. 1, January 1992, pp. 96-105.

[5] B. Hassibi, An Efficient Square-Root Algorithm for BLAST, manuscript, Bell Labs, 2000.

[6] B. Hassibi, A Fast Square-Root Implementation for BLAST, 34th Asilomar Conference on Signals, Systems, and Computers, pp. 1255-1259, 2000.

[7] A. Yasotharan, Multirate Sucessive-Cancellation Tx-Rx Design for MIMO Channel under ZF and BER Constraints, Canadian Conference on Electrical and Computer Engineering, May 2006, Ottawa, pp. 944-949.

[8] G. W. Stewart, Stochastic Perturbation Theory, SIAM Review, Vol. 32, No. 4, December 1990, pp. 579-610.

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd edition, The Johns Hopkins University Press, 1996.

[11] J. W. Brewer, Kronecker Products and Matrix Calculus in System Theory, IEEE Transactions on Circuits and Systems, Vol. CAS-25, No. 9, September 1978, pp. 772-781.

[12] A. Graham, Kronecker Products and Matrix Calculus: With Applications, Wiley, 1981.

[13] J. R. Magnus and H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, 2nd Edition, John Wiley & Sons, 1999.

[14] L. L. Scharf, Statistical Signal Processing - Detection, Estimation, and Time Series Analysis, Addison-Wesley, 1991.

[15] D. H. Brandwood, A complex gradient operator and its application in adaptive array theory, Proc. IEE, Special issue on adaptive arrays, vol. 130, Part F. pp. 11-17, February 1983.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) Defence R&D Canada – Ottawa 3701 Carling Avenue, Ottawa, Ontario, Canada K1A 0Z4	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

A simple expression for the matrix gradient of a diagonal element of R in QR decomposition

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used.)

Yasotharan, A.

5. DATE OF PUBLICATION (Month and year of publication of document.) December 2010	6a. NO. OF PAGES (Total containing information. Include Annexes, Appendices, etc.) 38	6b. NO. OF REFS (Total cited in document.) 15

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Memorandum

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – Ottawa
3701 Carling Avenue, Ottawa, Ontario, Canada K1A 0Z4

9a. PROJECT NO. (The applicable research and development project number under which the document was written. Please specify whether project or grant.) 15en01	9b. GRANT OR CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.) DRDC Ottawa TM 2010-247	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

(X) Unlimited distribution
(   ) Defence departments and defence contractors; further distribution only as approved
(   ) Defence departments and Canadian defence contractors; further distribution only as approved
(   ) Government departments and agencies; further distribution only as approved
(   ) Defence departments; further distribution only as approved
(   ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11)) is possible, a wider announcement audience may be selected.)

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

The QR matrix decomposition (QRD), or factorization, has many applications. In matrix computations, it is used to solve linear equations and least squares problems. In signal processing, it is used for adaptive filtering, adaptive beamforming/ interference nulling, and direction finding. In communications, it is used for adaptive equalization and transceiver design for multiple-input multiple-output (MIMO) channels.

When QRD is used in signal processing and communications, it is of interest to know the effects of noise. This work is a first step towards that goal.

Given matrix $\mathbf{A}$ with full column-rank M, we consider the unique decomposition $\mathbf{A} = \mathbf{QR}$ where $\mathbf{Q}$ is a matrix with M orthonormal columns and $\mathbf{R}$ is an $M \times M$ upper triangular matrix with real positive diagonal elements $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_M$. Treating $\hat{r}_i$ as a function of the elements of $\mathbf{A}$, a simple expression is derived for its matrix gradient with respect to $\mathbf{A}$.

Future work will aim to derive expressions for the gradients of all elements of $\mathbf{Q}$ and $\mathbf{R}$, and use these expressions to evaluate the effect of noise perturbation in $\mathbf{A}$. The present result is useful in optimizing certain MIMO decision-feedback communication systems.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Matrix Decomposition, Matrix Factorization, QR Decomposition, Unitary-Triangular Decomposition, Gram-Schmidt Orthogonalization, Orthogonal Projection, Matrix Gradient, Derivative w.r.t. Matrix, Multiple-input Multiple-output Communications, MIMO

Defence R&D Canada

Canada's leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE R&D DÉFENSE

www.drdc-rddc.gc.ca